Dundee Discussion Papers in Economics 302

Allanson, Paul; Cookson, Richard

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

UNIVERSITY OF DUNDEE

Dundee Discussion Papers in Economics

sire

Measuring healthcare quality variation using multicategory ordinal data: an application to primary care services in England

Paul Allanson & Richard Cookson

Economic Studies
University of Dundee
Dundee
DD1 4HN

Working Paper
No. 302
Corrected August 2021
ISSN: 1473-236X

Measuring healthcare quality variation using multicategory ordinal data: an application to primary care services in England

Paul Allanson *

Economic Studies, University of Dundee School of Business, Scotland, United Kingdom

Richard Cookson

Centre for Health Economics, University of York, England, United Kingdom

Abstract: The paper proposes a framework for measuring both the comparative quality of a set of healthcare providers and the variation in quality between them. The measures are directly calculable using the multicategory response data increasingly available from patient experience surveys but are also well defined for standard cardinal quality indicators. Moreover, they are sensitive to the full distribution of quality scores for each provider, not just the mean nor the proportion meeting some binary quality threshold. We illustrate our approach by providing comparable estimates of the variation in the quality of primary care services in England using three different sources of publicly available, general practice-level information: multicategory response patient experience data, ordinal inspection ratings and cardinal clinical achievement scores. Our results reveal considerable variation in the quality of primary care services at both local and regional level: for example, a randomly chosen patient from the best practice in England had a 52.0 percentage point higher chance of reporting a better rather than worse experience than a patient from anywhere in the country, whereas one from the worst practice had a 61.4 percentage point lower chance. Weak correlation between the comparative quality indices calculated using the alternative quality indicators provides evidence that inspection ratings and clinical process indicators capture different aspects of GP quality than patient experience. Finally we investigate the impact of standardisation, reporting results based on both raw and indirectly standardised practice quality profiles, with the latter based on the estimation of a distribution regression model.

Keywords: healthcare variation; comparative quality evaluation; ordinal data, primary care services, England

JEL codes: D63; I14; I18

Acknowledgements: Richard Cookson is supported by the Wellcome Trust (Grant No. 205427/Z/16/Z). The views expressed are those of the authors and not the Wellcome Trust.

* Corresponding author. Email address: p.f.allanson@dundee.ac.uk.

Introduction

Variation in the quality of healthcare services is a major policy concern in many countries (Busse et al., 2019), with patients in England commonly said to face a 'postcode lottery' in which their choice of healthcare provider and hence the quality of care they can expect to receive is largely determined by where they live. The measurement of such variation between healthcare providers or geographical areas is a routine exercise for quantitative indicators of structure, process and outcome quality (Mainz, 2003) using summary statistics such as the extremal quotient, standard deviation, coefficient of variation and systematic component of variation (Ibanez et al., 2009). However, these summary statistics are only appropriate for quality indicators measured on a cardinal scale, such as staff to patient ratios, proportions of patients receiving indicated treatment and risk-adjusted mortality rates. Nowadays, cardinal quality indicators are increasingly being supplemented by multicategory response information from patient experience surveys in which, importantly, respondents are typically asked to assess their quality of care by choosing between one of several ranked categories (e.g. very poor, poor, OK, good, very good). For example, England initiated a national patient survey programme in 2001 (DeCourcy et al., 2012), with surveys now regularly conducted of patient experience in a range of primary and secondary care settings (NHS England, 2021, https://www.england.nhs.uk/statistics/statistical-work-areas/patient-surveys/).

A critical limitation of this patient-reported data for the summary evaluation of both the performance of individual healthcare providers and the variation between them is its qualitative or ordinal nature. In particular, the mean is not well defined for polytomous categorical response data, which in turn severely restricts the choice of dispersion measures. A common workaround has been to impose some numerical scale on the ordinal data, but the resultant ranking of healthcare

providers by mean quality levels will not in general be robust to simple monotonic transformations of the chosen scale (cf. Bond and Lang, 2019) and this non-robustness problem extends to measures of variation that are a function of the mean (Allison and Foster, 1994). Another popular option is to collapse the number of categories to yield a binary 0/1 indicator that is amenable to analysis in terms of the proportion of patients reporting good (as opposed to not good) care (see e.g. Bruyneel et al., 2017), but the choice of cutoff is arbitrary and information is inevitably discarded in the process. Neither of these standard approaches therefore is particularly satisfactory despite their widespread use in practice.

The main contribution of this paper is to propose a measurement framework that is directly applicable to both ordinal and cardinal quality indicators without the need to first convert them to some common metric. More specifically, our framework makes use of ordinal information about the care quality profiles or distributions of all healthcare providers serving some population of interest to provide intelligible, patient-oriented measures of both the comparative quality of each provider and the variation in quality between them. The comparative quality of a provider is defined as the difference in the chances that the quality of care received by a randomly chosen patient treated by that provider will be better rather than worse than that received by a randomly chosen patient from the population as a whole. The measure of variation is equal to the average absolute difference in the chances that the quality of care received by patients will be better rather than worse as a result of being treated by one provider rather than another, leading us to call it the 'lottery' index. This index will take a minimum value of zero if all quality profiles are identical such that there is no difference in the chances that a randomly chosen patient treated by one provider will receive better rather than worse care than one treated by another. Conversely, it will take a maximum value of one if the quality of care provided by any one provider is certain to be

either strictly better or strictly worse than that provided by any other, which will only be the case for non-overlapping quality profiles. The intuition and mathematics behind our measures are set out in detail in the conceptual framework section below.

We show how our measurement framework can generate useful new insights by applying it to three different practice-level indicators of the quality of primary care services in England – categorical response data from the annual GP Patient Survey (GPPS), ordinal inspection ratings from the Care Quality Commission (CQC), and cardinal measures of process quality from the Quality and Outcomes Framework (QOF) – all of which are published in searchable online databases to help inform patients' choices. Primary care services in England are delivered through general practices ('practices' hereafter) with the average practice responsible for the care of about 7000 adult patients. The CQC, the independent regulator of health and social care service providers, reported wide variation between practices in the mean number of full-time equivalent (FTE) general practitioners (GPs) per head of registered population in 2018/19, with the geographical concentration of poor quality care, as shown by inspection ratings, making it difficult for people living in some areas to access good care (CQC 2019a, pp19-20). All practices are a member of one of nearly 200 Clinical Commissioning Groups (CCGs), which are responsible for the planning and commissioning of health care services for their local area. NHS England and Ipsos MORI (2019, p.10) report considerable variation across individual CCGs in the proportion of patients describing their practice as either fairly or very good in the 2019 GPPS, ranging from 69.1% to 92.1%. Patients were given the right to choose their practice in 2015, with the aim of improving the quality of access to GP services, although practices are not bound to accept patients living outside their catchment area. Santos et al. (2017) investigate patients' choice of family doctor and show that individuals are more likely to choose practices with higher standards of care

as measured by their total QOF score across all achievement indicators, trading off practice quality against distance.

Policy concern about variation in the quality of healthcare services relates specifically to that part of the variation not warranted by differences in patient need or preferences. Accordingly, measures of healthcare performance are often standardised with the aim of identifying this unwarranted variation by controlling for the effects of differences in patient characteristics not under the control of providers such as age, sex and ethnicity (see e.g. Public Health England, 2015; Dartmouth Atlas of Healthcare). To investigate the impact of standardisation on variation in primary care quality we report results based on both raw and indirectly standardised practice quality profiles, where the latter are what would be expected if quality outcomes conditional upon demographic characteristics were the same in each practice as in England as a whole.

The main empirical analysis is based on data from the 2019 English GPPS questionnaire, which was sent out to more than 2 million people asking for feedback on their experiences. Practice-level experience data, weighted by age and gender to resemble the population of eligible patients within each practice, are reported for nearly 7000 practices across 195 CCGs. We make use of the data on the proportions of patients in each practice reporting their overall experience as very poor, fairly poor, neither good nor poor, fairly good, and very good to explore the variation in primary care quality both between practices within each CCG and between CCGs in England. We also investigate the variation in primary care quality between CCGs using the CQC overall rating and total QOF score for each practice to see if these indicators provide ordinally equivalent information to the GPPS on some common latent 'primary care quality' characteristic.

Comparative quality indices are also calculated for all practices and CCGs using the GPPS data, and for all CCGs using the CQC and QOF data.[1]

The remainder of the paper is organized as follows. The next section introduces the conceptual framework, motivating the definitions of the comparative quality and lottery indices and outlining the indirect standardisation procedure. Section 3 discusses the various sources of data on practice quality which are employed in the empirical study, with the results presented in Section 4. The final section provides a discussion of the findings and concludes.

2. Conceptual Framework

The basic building block of our measurement framework is the comparative evaluation of the quality profiles of pairs of healthcare providers (i.e. practices or CCGs) based on information about the care quality profile or distribution of each healthcare provider. We start with a simple numerical example to provide the intuition behind the approach, before turning to the general mathematical formulation and properties of the comparative quality and lottery indices. Finally, we outline the indirect standardisation procedure.

[1] Analysis of the CQC and QOF data is restricted to the CCG level because the practice-level quality profiles for these indicators are degenerate distributions, consisting simply of an overall rating or score. This does not prevent the application of the measurement framework at the practice level but it does limit the informational value of the resultant indices. In particular, for a continuous quality indicator the comparative quality indices of practices will simply be given by their rank in the population-level quality distribution less half, while the between-practice lottery index will equal one in the absence of ties.

2.1 Measuring pairwise quality differences

Figure 1 provides an example in which the quality profiles for two practices, A and B, are given as the proportion of patients in each practice who report their care as either 'poor', 'OK', or 'good' – a three-valued ordinal scale. We first note that neither conversion to a numerical scale nor dichotomisation of the categories leads to a robust ranking of the quality profiles of the two practices. With numerical scaling, the mean quality of the two practices will be the same if the response options are assumed to be evenly spaced, being equal to 2.1 if the categories are scored 1, 2 and 3. But A has the higher mean if the distance between good and OK is greater than between OK and poor, whereas B has the higher mean if the opposite is the case. With dichotomisation, A has the higher proportion of patients reporting quality as good (rather than OK or poor) but a lower proportion reporting quality as either good or OK (rather than poor). It follows that neither approach can provide a robust basis for an analysis of the variation in quality between practices.

Figure 1: Two practice lottery example

		Practice B		
	Quality Profile	Good **40%**	OK **30%**	Poor **30%**
Practice A	Good **50%**	**20%** *Both good*	**15%** *A good B OK*	**15%** *A good B poor*
	OK **10%**	**4%** *B good A OK*	**3%** *Both OK*	**3%** *A OK B poor*
	Poor **40%**	**16%** *B good A poor*	**12%** *B OK A poor*	**12%** *Both poor*

Note: The light and dark grey shaded boxes show respectively the proportion of draws in which care is better in A than B and vice versa, assuming care quality for each practice is chosen independently and at random from the quality profile for that practice.

The calculation of the lottery index may be thought of in terms of the outcome of a lottery in which the patient has an equal chance of being assigned to A or B with the quality level for each practice determined by a random draw from the quality profile for that practice. The patient 'wins' or 'loses' depending on whether they are assigned to the practice with the higher or lower randomly chosen quality level, and will be indifferent to the lottery outcome if the quality levels delivered by the two practices are the same. Patients have a (15+15+3)=33% chance of 'winning' if assigned to A, a (4+16+12)=32% chance of 'winning' if assigned to B and will be indifferent to the lottery outcome in the remaining (20+3+12)=35% of draws. Hence the difference in 'winning' chances of (33-32) = 1% provides a measure of the degree to which the profile of A is superior to that of B. We proceed to calculate the lottery index as the absolute value of this difference, where this is equal by definition to the absolute difference in the chances that a patient randomly assigned to one practice will receive better rather than worse care than if assigned to the other.

More generally, consider some population in which each individual is a patient of one (and only one) of a set of $K \geq 2$ healthcare providers, such that the patient list of each provider is independent of that of any other. Let $\mathrm{P}(Q_k \geq Q_{k'}) = \mathrm{P}(Q_k > Q_{k'}) + \mathrm{P}(Q_k = Q_{k'})$ be the probability that the quality of care received by a randomly chosen patient with provider $k \in K$ is at least as good as − i.e. strictly better than or the same as − that received by a randomly chosen patient with provider $k' \in K$. The *pairwise quality difference* is defined as the difference in the chances that the quality of care received by a randomly chosen patient with provider k' is (strictly) better rather than worse than that received by one with provider k:

$$\Delta_{kk'} = -\Delta_{k'k} = \mathrm{P}\left(Q_{k'} \geq Q_k\right) - \mathrm{P}\left(Q_k \geq Q_{k'}\right) = \mathrm{P}\left(Q_{k'} > Q_k\right) - \mathrm{P}\left(Q_k > Q_{k'}\right); \qquad \forall k, k' \in K \tag{1}$$

$\Delta_{kk'}$ will take a value of zero if the quality profiles of the two providers are equivalent, although this does not necessarily imply that they are identical; a maximum value of one when the worst quality of care provided by provider k' is strictly better than the best quality provided by provider k; and a minimum value of minus one when the opposite is the case.

The normative significance of the pairwise quality difference derives from the use of the statistical preference criterion (De Schuymer et al., 2003) for the comparative evaluation of quality profiles. According to this criterion one profile is better than another if the patient receiving the (strictly) higher quality care of any randomly chosen pair of patients is more likely to be registered with the first rather than the second provider. The criterion is more general and powerful than first-order stochastic or rank dominance (De Baets and De Mayer, 2007), which is commonly employed to compare ordinal distributions but can lead to incomplete orderings. Statistical preference will always say whether one quality profile is better, worse or equivalent to another, whereas rank dominance often leaves things undefined – neither better nor worse, but not equivalent either. Thus, A and B in the numerical example are not comparable by rank dominance since the proportion of patients who receive poor care is lower in B but the proportion receiving no better than OK care is lower in A. Moreover, statistical preference is not only able to rank all quality profiles but also provides a 'graded' comparison of them (De Baets and De Mayer, 2007), with the pairwise difference in winning chances offering a readily intelligible measure of the degree to which one profile is better or worse than another.

The comparative quality index

A summary measure of comparative quality for each provider can be obtained by calculating a pairwise index for it relative to some common benchmark patient quality profile, such as that of the whole population (Allanson, 2021). The *comparative quality index*:

$$\Delta_k = \sum_{k'=1}^{K} p_{k'} \left(\mathrm{P}\left(Q_k > Q_{k'}\right) - \mathrm{P}\left(Q_{k'} > Q_k\right) \right) = \sum_{k'=1}^{K} p_{k'} \Delta_{k'k}; \qquad \forall k \in K \tag{2}$$

offers a summary measure of the quality of provider *k* compared to all *K* providers, where $p_{k'}$ is the proportion of total registrations with provider *k′*. The index may be used to generate a complete ranking of providers by quality but will generally be more informative than a simple measure of 'league table' position: Δ_k can take values in the closed interval from $-(1-p_k)$ to $+(1-p_k)$, with the sign of the index indicating whether the care quality of organization *k* is better or worse than the benchmark and its magnitude indicating the scale of any difference. By construction, Δ_k takes a weighted average of zero across all providers, i.e. $\sum_k p_k \Delta_k = 0$.

The lottery index

The lottery index is simply the average absolute value of the pairwise quality difference $\left|\Delta_{kk'}\right|$ over all distinct pairs of providers. Mathematically, it is defined as a normalized version of the Allanson (2021) headcount stratification index:

$$L = \left(\sum_{k=1}^{K} \sum_{k'=1}^{K} p_k p_{k'} \left|\Delta_{kk'}\right| \right) \Bigg/ \left(1 - \sum_{k=1}^{K} p_k^2 \right) \tag{3}$$

where the normalization factor $\left(1-\sum_k p_k^2\right)$ implies that *L* may be interpreted as the patient-weighted mean absolute difference in the chances that quality will be better rather than worse as a result of being cared for by one provider rather than another. The interpretation in terms of the average absolute difference in the chances of winning rather than losing over all possible pairwise lotteries following directly from the definition of the pairwise index $\left|\Delta_{kk'}\right|$.

Alternatively, the index may be interpreted as a measure of the potential value to patients of exercising the right to choose their healthcare provider rather than it being determined by the accident of where they live. This follows since $\left|\Delta_{kk'}\right|$ in (3) may also be written as:

$$\left|\Delta_{kk'}\right| = \left(2\max\left(\mathrm{P}\left(Q_{k'} > Q_k\right), \mathrm{P}\left(Q_k > Q_{k'}\right)\right) - \left(\mathrm{P}\left(Q_{k'} > Q_k\right) + \mathrm{P}\left(Q_k > Q_{k'}\right)\right)\right); \qquad \forall k, k' \in K \tag{4}$$

so L may also be interpreted as twice the mean increase in the probability that patient care will be better than it would otherwise have been if patients chose the provider with the better quality profile of any pair of providers rather than being randomly assigned to one of them.

A third interpretation is in terms of the degree of "postcode discrimination" faced by patients on the basis of where they live due to the variation in care quality across providers. Specifically, L may be interpreted as a summary measure of discrimination between pairs of providers given that $\Delta_{kk'}$ is formally equivalent to the Le Breton *et al.* (2008) first-order discrimination index Δ_1 if provider k' has the better profile of the two providers.

L will take a minimum value of zero if and only if the comparative quality − but not necessarily the quality profiles − of all providers is the same and a maximum value of one if there is complete separation of the patient lists for each provider into disjoint strata in the population quality profile. The index is sensitive to any change in the quality of care received by any patient unless the change is over some quality range occupied exclusively by others cared for by the same provider as the patient. For binary 0/1 quality indicators (e.g. good or bad), L is simply the weighted average absolute of the pairwise differences in the proportion of patients receiving good care. But, as shown by the example, it can also be calculated for ordinal measures with three or more categories without the need for dichotomisation.

Given independent patient lists, the simplest way to compute L for an ordinal quality indicator is to calculate the pairwise indices using the approach employed in the numerical example and then take the weighted average over all pairs. A more computationally efficient approach if there are more than 3 health categories makes use of the relation $\Delta_{kk'} = \left(1 - 2\left[\mathrm{P}(Q_k > Q_{k'}) + 0.5\,\mathrm{P}(Q_k = Q_{k'})\right]\right)$ in the first step. For cardinal indicators, the pairwise indices can be calculated exactly from the relation $\Delta_{kk'} = G_b / G_B$ if practice k' has the higher mean quality of the two providers (Monti and Santoro, 2011), where G_B is the conventional between-group Gini coefficient (Pyatt, 1976) and G_b and is the variant proposed in Yitzhaki and Lerman (1991). Alternatively, L may be approximated to any required degree of accuracy by rounding the data and then treating the resultant discretised variable like any other ordinal indicator.

Standardisation of practice quality profiles

Previous studies have revealed systematic differences in how patients from different demographic groups evaluate the quality of primary care services (see e.g. Paddison et al., 2012; Lyratzopoulos et al., 2012). Individual response data from the GPPS could in principle be used to estimate directly standardised quality profiles calculated on the basis that all practices had the same demographic composition as the whole population. However, the sample size of the GPPS is not large enough to provide reliable estimates of demographic-specific quality profiles at the practice level and the approach is in any case inapplicable to the practice-wide CQC ratings and QOF scores. Instead we employ an indirect standardisation procedure based on the estimation of a distribution regression model (Chernozhukov et al., 2013) for each quality indicator to predict the practice quality profiles that would be expected if quality outcomes conditional upon demographic characteristics were the same in each practice as in England as a whole. Specifically, the proportion of the patients of a

practice expected to experience a quality level no better than q ($q = 1, \ldots. Q-1$ of Q discrete quality levels) is given by the prediction from a binary choice model in which the dependent variable takes a value equal to the proportion of patients reporting experience no better than q.

Data and Methods

Data

Patient experience data for 6926 practices were obtained from the 2019 results of the annual GPPS (NHS England, 2019). The survey asked patients about a range of issues associated with using the services offered by their practice, including how they would describe their overall experience using a 5-category semantic differential scale, as well as various questions about their own personal circumstances. The specific question was: "Overall, how would you describe your experience of your GP practice?", with response categories: "Very good", "Fairly good", "Neither good nor poor", "Fairly poor", "Very poor". Postal questionnaires were sent out in January 2019 to 2.33 million adult patients in England of whom 770512 in 6999 practices completed the survey representing a response rate of 33.1% (Ipsos MORI, 2019). All practices listed on NHS Digital as having eligible patients were included in the survey apart from an unspecified number that chose to opt out as they felt it was inappropriate to their patient population. Patients were eligible for inclusion in the survey if they had a valid NHS number, had been registered with a practice continuously for at least six months before being selected, and were 16 years of age or over. The sample was based on a proportionately stratified, unclustered design, with the sample size for each practice selected to ensure that confidence intervals were as consistent as possible between practices. Practice-level data are published on a weighted basis to ensure that the results are more representative of the population of adult patients registered with each practice by correcting for the sampling design and to reduce the impact of non-response bias. No overall experience data are

available for 73 practices due to the suppression of data for questions answered by fewer than 10 people to protect confidentiality.

Inspection ratings data for 6670 practices was obtained from the January 2019 CQC Care Directory (CQC, 2019b). The Care Directory is updated monthly and includes the latest published ratings of all practices that have been subject to inspection in England, which in January 2019 dated back as far as November 2014. Practices are given an overall rating for the 'whole population' of service users on a 4-category semantic differential scale following a visit by an inspection team and taking account of the views of both patients and staff. The overall rating is based on a detailed assessment of the quality of care across six patient subgroups in terms of whether the service is safe, effective, caring, responsive to people's needs and well-led. The most recent rating was used for practices with multiple ratings based on different inspection dates. The rating for the main branch of a practice was used where ratings were available for more than one location.

QOF scores for 6854 practices with achievement data were obtained from the QOF 2018-19 results (NHS Digital, 2019a). The QOF is a voluntary, annual incentive payment scheme for all practices in England that rewards practices for the provision of 'quality care', with 95.1% of practices participating in the reporting year 1 April 2018 to 31 March 2019. The QOF provides an indication of overall practice achievement through a points system, with points awarded against a range of 77 clinical care and public health indicators based, for example, on the proportion of patients on specified disease registers who receive defined interventions. The headline measure of practice achievement published by NHS England is percentage attainment of the maximum 559 QOF points available, but an alternative measure is also provided which takes account of instances where practices cannot achieve points because they have no patients pertinent to an indicator. We

use the publicly reported scores and refrain from making an adjustment by adding 'exception reported' patients back into the population denominator, which typically provides a less favourable measure of performance. QOF percentage attainment data are rounded to 1 decimal place to calculate the indirectly standardised quality profiles.

Methods

The main analysis of patient experience data was based on the full GPPS sample of 6,926 practices. A sub-set of 6,427 matched practices with valid GPPS, CQC and QOF data was used to generate comparable CCG-level results for all three practice quality indicators. All sample practices belonged to one of 195 CCGs, with the number per CCG varying between 10 and 169, and a mean of 35.5. Practice weights based on the number of registered patients aged 16 years old and over in December 2018 (NHS Digital, 2019b) were used to construct CCG quality profiles as weighted averages of sample practice profiles and, after adjusting for missing practices within each CCG, to ensure the national representativeness of results at the CCG level. Practice-level comparative quality and within-CCG lottery indices were calculated using the GPPS practice quality profiles, and CCG-level comparative quality and between-CCG lottery indices using the CCG quality profiles for all three indicators.

We report both total and indirectly standardised indices. For the estimation of indirectly standardised quality profiles, distribution regression models for each practice quality measure were specified as a function of the sex, age group (16-24, 25-34, 35-44, 45-54, 65-74, 75-84, 85+) and ethnic (White, Asian, Black, Mixed, Other) composition of each practice patient list as reported in the GPPS data. The specifications also allowed for CCG-specific fixed effects, with predictions based on practice patient list composition and CCG patient population shares. In our base case analysis we employ a linear probability distribution regression model (LPDRM) for convenience

but, as a robustness check, also calculate indirectly standardised profiles using a generalized linear distribution regression model (GLDRM) with a probit link function and a binomial distribution with the parameter *n* set equal to the number of survey responses in a practice for the GGPS data, and to one for the CQC and QOF data. Estimated counterfactual cumulative proportions were censored where necessary to lie in the unit interval, with the resultant set of predictions scaled to match the sample mean. Finally, bootstrap standard errors were obtained for all comparative quality and lottery indices by the resampling of practices within each CCG to reflect the organizational structure. All analysis was conducted using Stata version 15.1.

Results

We first present results based on the full sample of practices with GPPS patient experience data, looking in turn at the indices calculated using the practice and CCG-level quality profiles. We subsequently compare the indices calculated using the GPPS, CQC and QOF CCG-level quality profiles constructed from the matched sample of practices with valid data for all three indicators.

GPPS patient experience

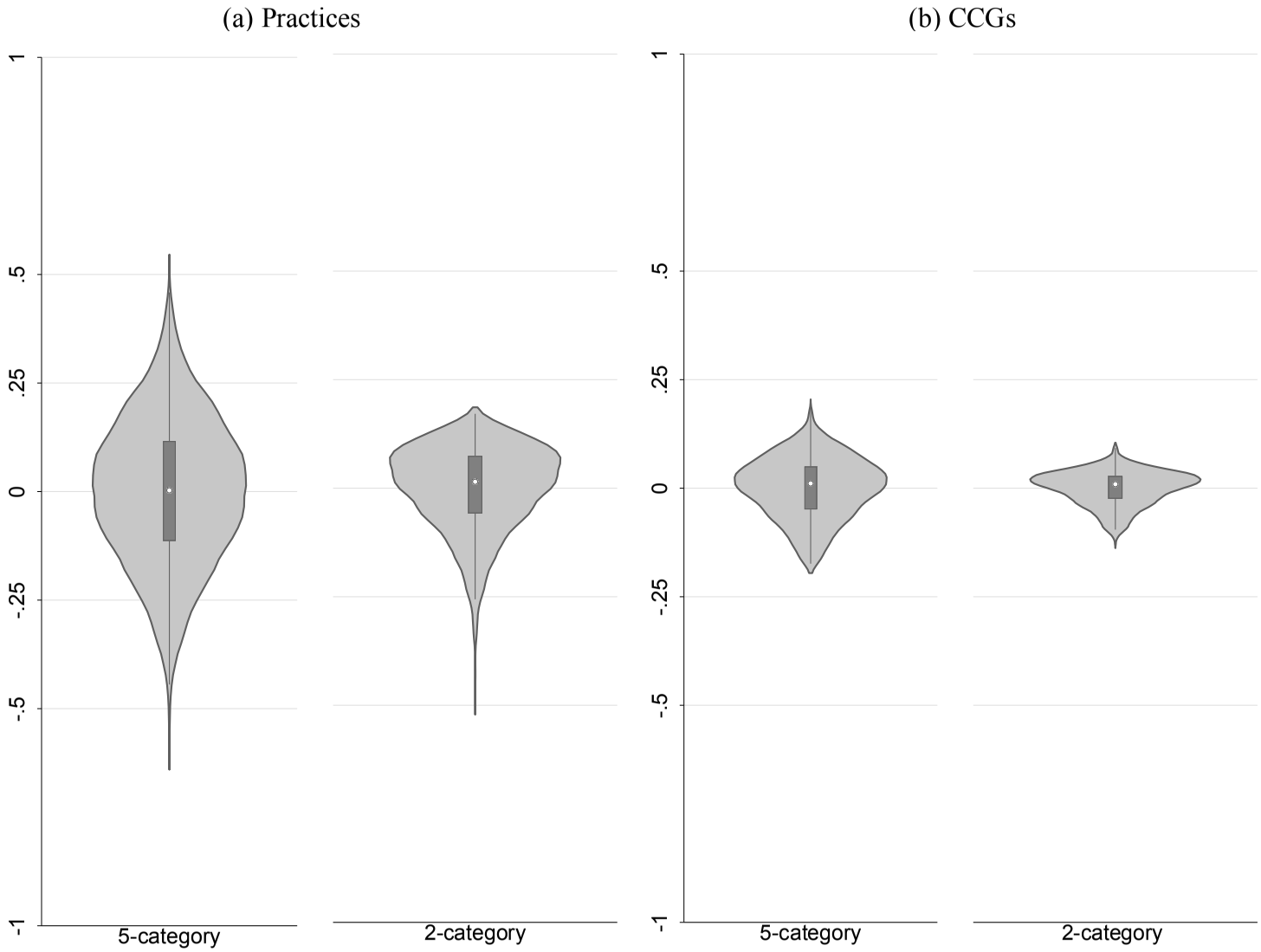
This section reports results based on the full GPPS sample of 6926 practices. The proportions of adult patients in England reporting their overall experience of their practice as very poor, fairly poor, neither good nor poor, fairly good and very good were 2.1%, 4.4%, 10.6%, 37.8% and 45.1% respectively. Scoring these responses 1 to 5, practice quality was 4.19 on average with a standard deviation of 0.30 across all practices. It might thus appear that the variation in reported experience between practices was low relative to the mean, but the coefficient of variation can be made arbitrarily large or small through the choice of alternative scoring schemes. For example, if the responses were scored instead from −2 to +2, with 0 providing a natural measure of neither good

nor poor, then the coefficient of variation would be 24.9% not 7.1%. Some other approach is therefore required to meaningfully assess the degree of variation in reported experience.

Figure 2(a) shows the distribution of practice-level comparative quality index values, which have a patient-weighted mean of zero by construction. The variation in comparative quality across individual practices is considerable, ranging from a 0.520 or 52.0 percentage point (pp) higher chance that a patient from the best practice would have reported a better rather than worse experience than one from anywhere in England to a 61.4pp lower chance for the worst practice. The standard deviation of the comparative quality index is 16.8pp, with within-CCG differences accounting for 83.0% of the variance in practice-level comparative quality and only 17.0% due to between-CCG differences. Thus there was much more variation between practices within each CCG than between CCGs, where the former is of more relevance for the exercise of patient choice given the evidence that patients are only willing to travel a limited distance to access better quality GP services (Santos et al., 2017).

Responses to the patient experience question are commonly collapsed into a dichotomous variable for presentational purpose by combining very poor/fairly poor/neither good nor poor into one category and fairly good/very good into the other (see e.g. NHS England and Ipsos MORI, 2019). However, the use of this binary quality indicator leads to a marked reduction in the ability to discriminate between ‘average’ and ‘good’ practices, while continuing to capture the extent to which ‘bad’ practices offer poorer quality care. Thus, a patient from the best practice is now estimated to have had only a 17.1pp higher chance of reporting a better rather than worse experience than one from anywhere in England, whereas a patient from the worst practice would have had a 50.7pp lower chance. Overall, dichotomisation leads to a substantial underestimate of the variation in the quality of care between practices with the standard deviation of the comparative quality index falling to 9.8pp as a result.

Figure 2. Violin plots of comparative GPPS patient experience indices for the whole of England:



The first row of results in Table 1(a) reports an average 17.8pp absolute difference in the chances that patient experience was better rather than worse as a result of being registered with one practice rather than another within the same CCG. Thus, on average, it was 8.9pp (=17.8/2) more likely that patient experience would have been better than it would otherwise have been as a result of being able to choose the better of any pair of practices within a CCG rather than being randomly assigned to one of them. Figure 3(a) maps the distribution of within-CCG lottery index values, ranging from a 9.0pp absolute difference in patients' chances of reporting a better rather than worse experience as a result of being registered with one practice rather than another in the most homogeneous CCG to a 30.3pp difference in the least. We note that the expected value of this index is not a function of the number of practices within a CCG although, unsurprisingly, the conditional variance is decreasing in the number of practices. Figure 3(b) further shows that the heterogeneity of practices within individual CCGs in terms of their demographic composition does not account for that much of the total variation in practice quality within CCGs, with predicted within-CCG variation highest in the more diverse and segregated metropolitan areas based on the LPDRM estimates in Table 2. The within-CCG lottery index based on the indirectly standardised profile was 4.6pp rather than 17.8pp, leaving a residual or 'unexplained' 13.1 pp average absolute difference in the chances that reported patient experience would have been better rather than worse as a result of being registered with one practice within a CCG rather than another.

Figure 2(b) shows that there was also considerable variation across individual CCGs in comparative quality levels, ranging from a 18.3pp higher chance that a patient from the best CGG would have reported better rather than worse experience than one from anywhere in England to a 17.4pp lower chance for the worst CCG. Dichotomisation again leads to a reduction in measured variation, particularly between 'average' and 'good' CCGs: the range in chances shrinks to 9.3 pp higher for the best CCG to 12.6 pp lower for the worst, that is to the difference in the proportion

Table 1. Lottery indices

		LPDRM results		*GLDRM results*	
	Raw or Unadjusted	Indirectly standardised	Residual	Indirectly standardised	Residual
(a) Full sample GPPS					
Average within-CCG indices					
GPPS 5-category	0.1775 **	0.0461 **	0.1314 **	0.0481 **	0.1294 **
	0.0015	*0.0022*	*0.0021*	*0.0023*	*0.0022*
GPPS 2-category	01002 **	0.0257 **	0.0745 **	0.0260 **	0.0742 **
	0.0010	*0.0012*	*0.0012*	*0.0013*	*0.0013*
Between-CCG indices					
GPPS 5-category	0.0789 **	0.0555 **	0.0234 **	0.0573 **	0.0216 **
	0.0027	*0.0026*	*0.0038*	*0.0028*	*0.0039*
GPPS 2-category	0.0433 **	0.0304 **	0.0129 **	0.0304 **	0.0130 **
	0.0016	*0.0016*	*0.0023*	*0.0016*	*0.0024*
(b) Common sample					
Between-CCG indices					
GPPS 5-category	0.0789 **	0.0548 **	0.0241 **	0.0562 **	0.0227 **
	0.0024	*0.0025*	*0.0039*	*0.0025*	*0.0038*
CQC 4-category	0.0986 **	0.0158 **	0.0827 **	0.0017 **	0.0962 **
	0.0064	*0.0041*	*0.0084*	*0.0002*	*0.0064*
QOF cardinal	0.2931 **	~	~	~	~
(% achievement 559)	*0.0088*				
– clinical domain	0.2875 **	~	~	~	~
	0.0084				
– public health domain	0.2259 **	~	~	~	~
	0.0087				
– public health AS	0.2313 **	~	~	~	~
	0.0077				
QOF cardinal	0.2931 **	~	~	~	~
(% achievement practice	*0.0088*				
QOF discretised	0.2925 **	0.0937 **	0.1987 **	0.1090 **	0.1835 **
	0.0091	*0.0102*	*0.0140*	*0.0117*	*0.0143*
QOF 5-category	0.2577 **	0.0843 **	0.1722 **	0.0919 **	0.1658 **
	0.0068	*0.0080*	*0.0109*	*0.0091*	*0.0117*

*'Residual' indices are calculated as the difference between the corresponding raw and indirectly standardised indices and reflect that part of the total variation in care quality not 'explained' by the relevant distribution regression model. Bootstrapped standard errors based on 50 replications are in italics. *p<0.05, **p<0.01. Source: Own calculations from GPPS, CQC and QOF data.*

Figure 3. Within-CCG lottery indices based on GPPS 5-category patient experience responses

(a) Raw or unadjusted profiles

(b) Indirectly standardised profiles

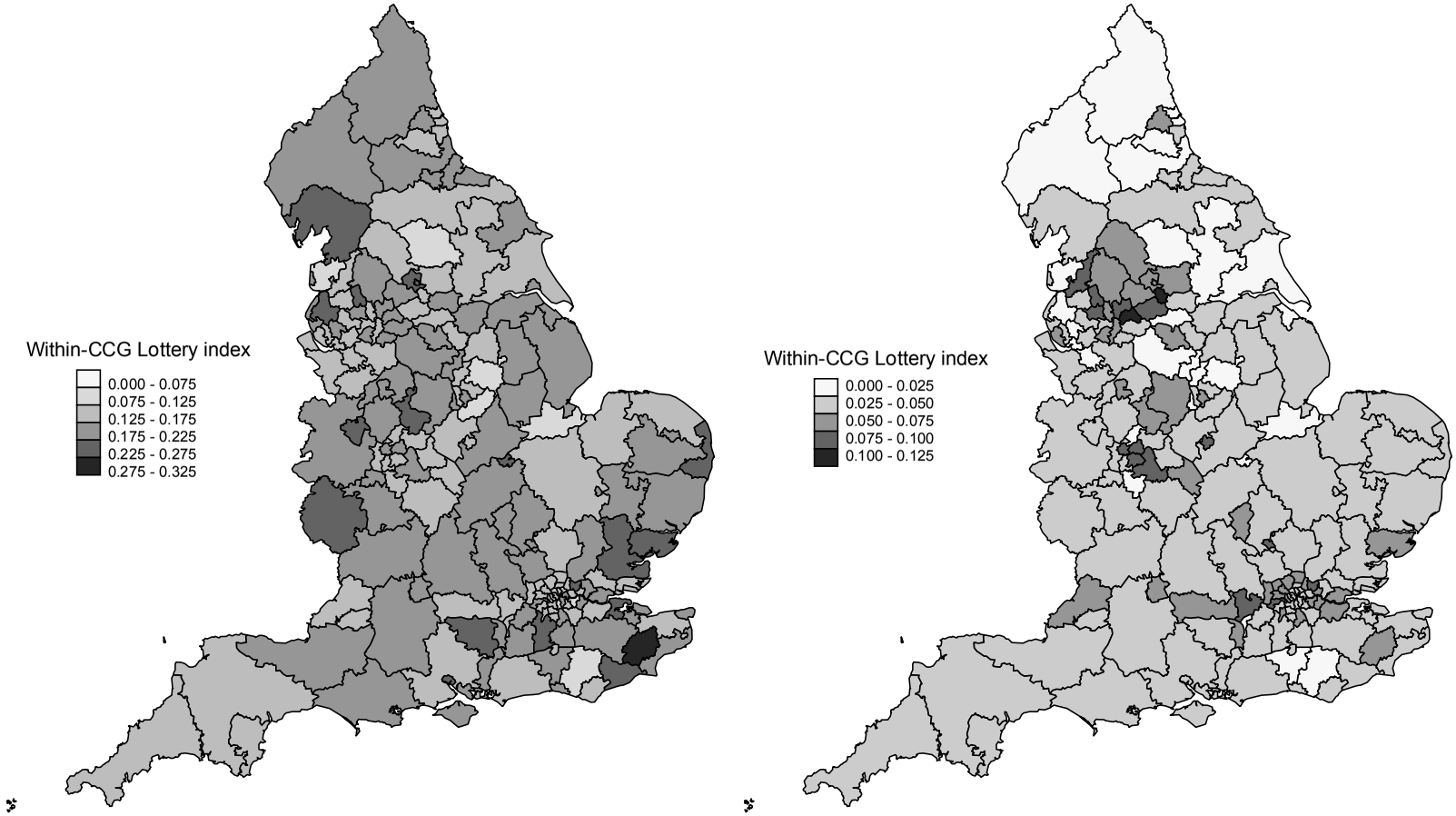


Table 2. Fixed effects estimates of the LPDRM: GPPS 5-category

	Reported experience no better than:			
	very poor	fairly poor	neither good nor poor	fairly good
Dependent variable	*P(q≤1)*	*P(q≤2)*	*P(q≤3)*	*I(q≤4)*
FEMALE	−0.0134 *	−0.0097	−0.0569*	−0.0210
	0.0067	*0.0153*	*0.0252*	*0.0323*
AGE 16-24	−0.0513 **	−0.0939**	−0.0906**	−0.0516
	0.0089	*0.0196*	*0.0351*	*0.0510*
AGE 25-34	−0.0186	−0.0372	−0.0122	0.0729
	0.0102	*0.0201*	*0.0343*	*0.0481*
AGE 35-44	−0.0179	−0.0271	−0.0132	0.0859
	0.0110	*0.0219*	*0.0358*	*0.0522*
AGE 55-64	−0.0295 *	−0.0590*	−0.0824	−0.0253
	0.0121	*0.0257*	*0.0436*	*0.0640*
AGE 65-74	−0.0464 **	−0.1217**	−0.2249**	−0.2903**
	0.0168	*0.0345*	*0.0593*	*0.0836*
AGE 75-84	−0.0631 **	−0.1618**	−0.1625*	−0.1777
	0.0207	*0.0512*	*0.0809*	*0.1060*
AGE 85+	−0.0666 *	−0.1460*	−0.2861**	−0.3246*
	0.0278	*0.0601*	*0.1058*	*0.1595*
BLACK	0.0291 **	0.0495**	0.0949**	0.1422**
	0.0080	*0.0147*	*0.0311*	*0.0425*
ASIAN	0.0416 **	0.0705**	0.1352**	0.1781**
	0.0034	*0.0080*	*0.0134*	*0.0178*
MIXED	0.0119	−0.0175	0.0620	0.1482
	0.0212	*0.0420*	*0.0712*	*0.1000*
OTHER	0.0192	0.0287	0.0617	0.0904
	0.0156	*0.0290*	*0.0515*	*0.0716*
constant	0.0492 **	0.1134**	0.2422**	0.5314**
	0.0079	*0.0162*	*0.0266*	*0.0381*
Practices	6926	6926	6926	6926
CCG clusters	195	195	195	195
R^2	0.186	0.186	0.223	0.229
RMSE	0.024	0.051	0.087	0.127

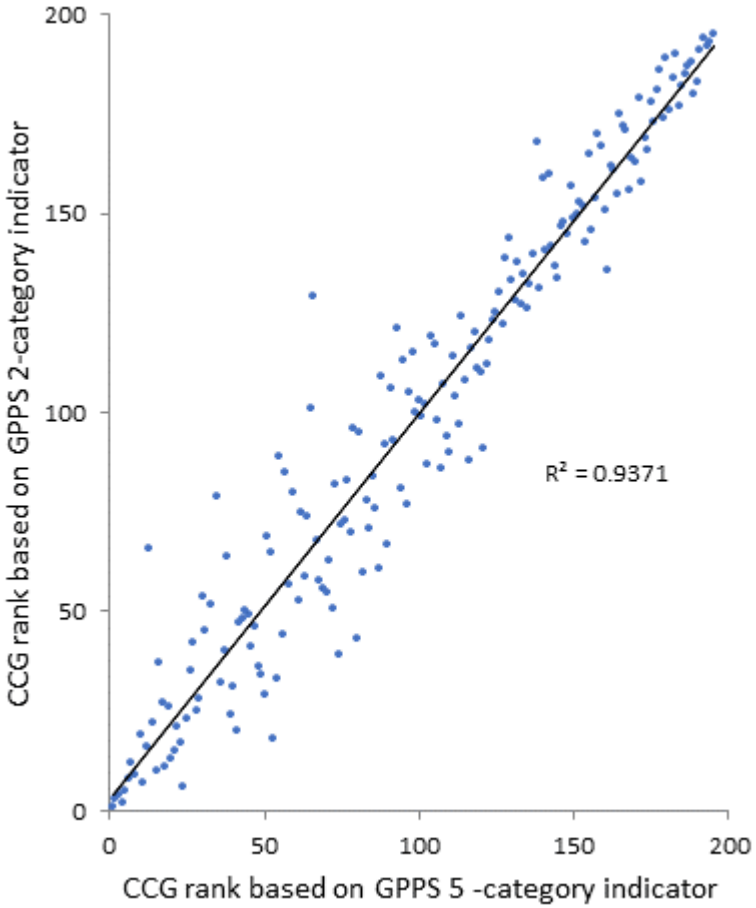
*The dependent variable P(q≤c) takes a value equal to the proportion of patients in a practice reporting their experience as no better than category c (c=1,2,3,4). Positive (negative) coefficients imply higher (lower) proportions than for the reference group of White men aged 45-54 in NHS Darlington CCG: for example, a 1pp increase in the proportion of Asian patients is predicted to lead to ceteris paribus increases of 0.0416 pp, 0.0705 pp, 0.1352pp and 0.1781pp in the proportions reporting experience no better than very poor, fairly poor, neither poor nor good, and fairly good respectively. CCG fixed effects not reported. Robust CCG-clustered standard errors are in italics. *p<0.05, **p<0.01. Source: Own calculations from GPPS data.*

of patients reporting their experience as either fairly or very good between the best and worst performing CCGs (cf. NHS England and Ipsos MORI, 2019, p.10). Dichotomisation also leads to changes in the ranking of CCGs as shown by Figure 4(a), with the Kendall's rank correlation coefficient τ_a between the two rankings implying that the full and dichotomised measures are 86.5 pp (95% CI, 0.837 to 0.893) more likely to agree than differ over which of any pair of CCGs had the better quality profile.

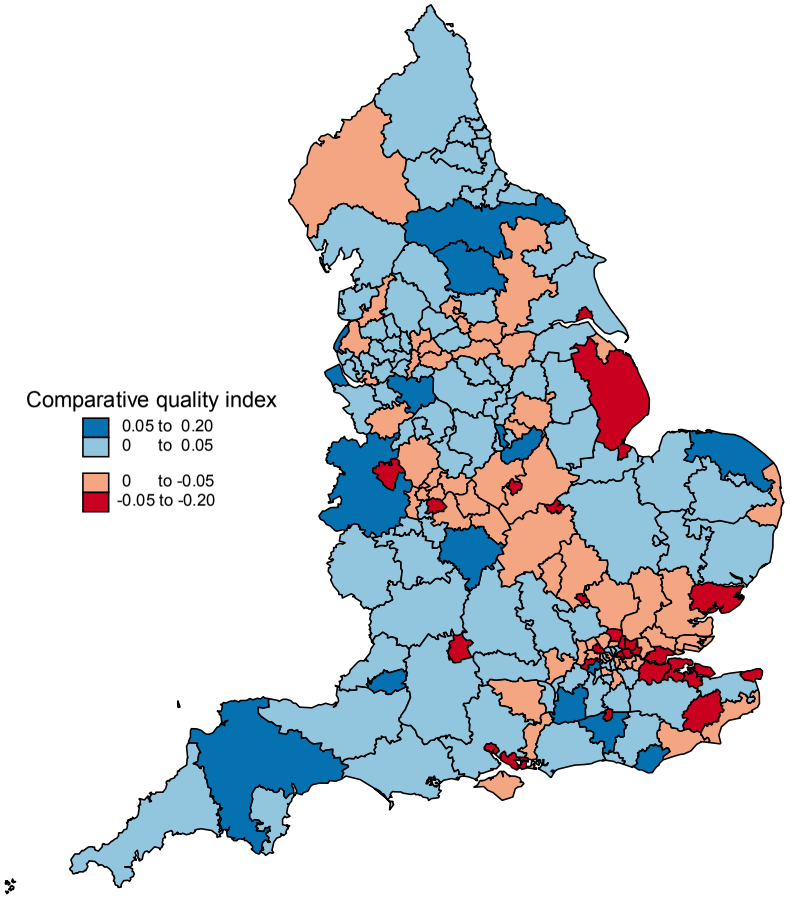
Figure 5(a) shows that comparative patient experience levels tend to be worse in metropolitan regions and surrounding areas than in the more rural 'shire' counties. Figure 5(b) shows that this geographical pattern is strongly associated with demographic differences between CCGs, with the LPDRM estimates in Table 2 implying that CCGs containing higher proportions of prime working age adults (25-54 year olds), Asians and Blacks were likely to have worse quality profiles than those with more young and older adults, and Whites. The between-CCG lottery index would have been 5.5pp rather than 7.9pp if the only source of variation in practice quality was differences in the demographic composition of patient lists, leaving an unexplained or residual 2.3pp absolute difference in the chances that patient experience was better rather than worse as a result of being registered with one CCG rather than another. Finally, Figure 4(b) shows that dichotomisation does not fundamentally change the underlying geographical pattern of comparative performance but does lead to a loss of contrast between better and worse performing CCGs.

Figure 4. CCG comparative quality indices based on dichotomised GPPS patient experience responses

(a) Scatterplot of CCG ranks against 5-category ranking



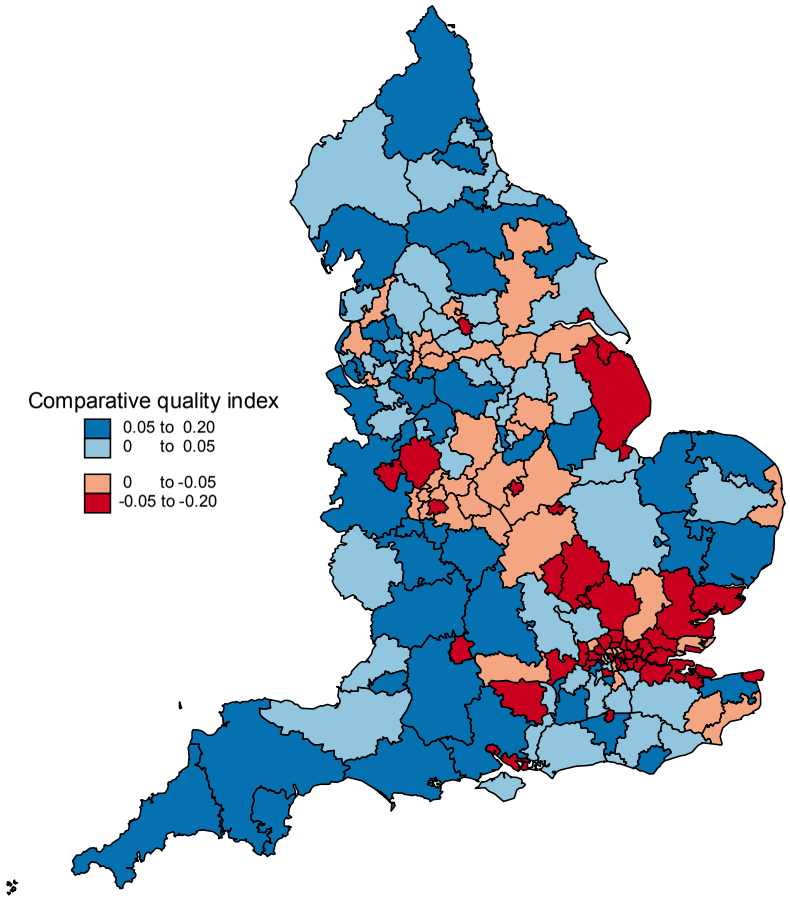
(b) Raw or unadjusted profiles



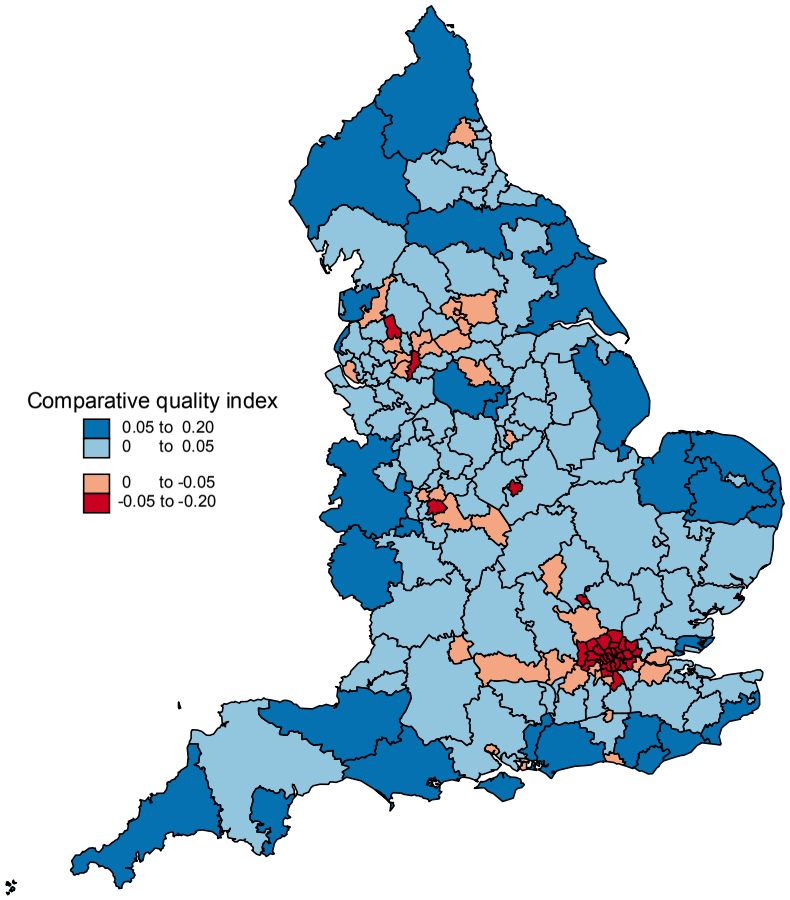
CCGs are ranked in descending order of comparative quality

Figure 5. CCG comparative quality indices based on GPPS 5-category patient experience responses

(a) Raw or unadjusted profiles



(b) Indirectly standardised profiles



Comparative analysis of three practice quality measures

This section reports results based on the matched sample of 6427 practices with valid GPPS, CQC and QOF quality data. The left-hand plot in Figure 6 and first column of Table 1(b) present results based on the GPPS CCG-level quality profiles, which are virtually the same as those discussed above for the full GPPS sample. We compare these results to those obtained with the CQC and QOF indicators.

The proportions of the patient population in England registered with a practice rated by the CQC as inadequate, requires improvement, good and outstanding were 0.7%, 2.9%, 90.4% and 6.0% respectively. Figure 6 also plots the CCG comparative quality indices based on CQC inspection ratings, ranging from a 74.7pp higher chance that a patient from the best CGG would have been in a practice with a higher rather than lower rating than one from anywhere in England to a 22.8pp lower chance for the worst CCG. The between-CCG lottery index of 9.9 ppts reported in Table 1B is nevertheless similar in magnitude to that for the GPPS measure. Figure 7(a) shows that the association between the ranking of CCGs by patient experience and inspection rating is weak. Kendall's τ_a is only 0.305 (95% CI, 0.203 to 0.406), implying that there was only a 30.5pp higher chance that the two measures would agree rather than differ over which of any pair of CCGs had the strictly better quality profile. The null hypothesis that τ_a is equal to 1, which would be the value if the two measures produced identical rankings of CCGs, can be rejected decisively implying that GPPS patient experience and CQC inspection rating data do not provide alternative sources of ordinally equivalent information on some common latent 'primary care quality' characteristic. Finally, the LPDRM indirectly standardised lottery index of 0.0158 is only 16% of the raw value, based on the distribution regression results in the Appendix, as very little of the variation in inspection ratings between CCGs can be accounted for by practice-level differences in demographic composition.

Figure 6. Violin plots of raw CCG comparative quality indices based on alternative measures.

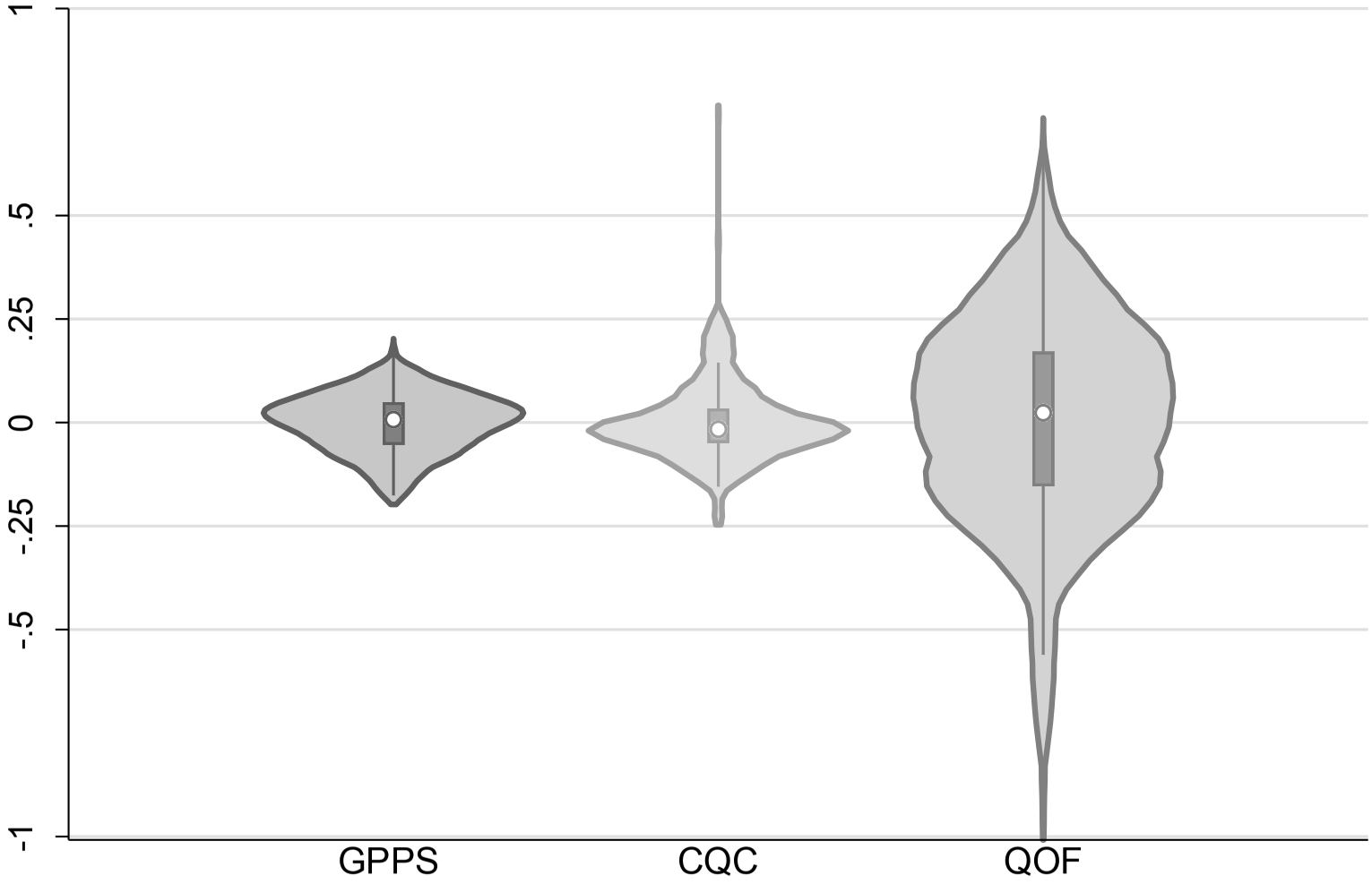
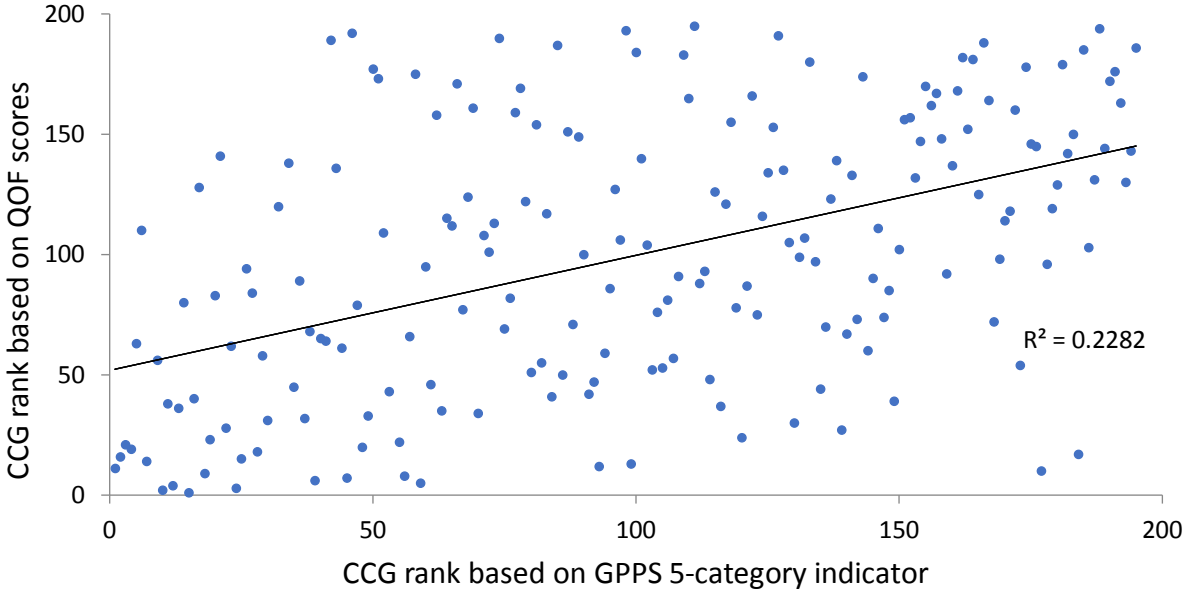


Figure 7(a). Scatterplot of CCG ranks by CQC ratings against GPPS 5-category ranking

CCGs are ranked in descending order of comparative quality

Figure 7(b). Scatterplot of CCG ranks by QOF score against GPPS 5-category ranking



CCGs are ranked in descending order of comparative quality

Levels of QOF achievement were very high with 14.5% of patients registered in practices achieving the maximum score of 559 QOF points, mean percentage achievement of 96.9 pp (541.6 points), and standard deviations of 5.4 pp (30.2 points) and 3.0 pp (16.7 points) at the practice and CCG levels respectively. The right hand plot of CCG comparative quality indices in Figure 6 is based on QOF scores, ranging from a 66.2pp higher chance that a patient from the best CGG would have been in a practice with a higher rather than lower QOF score than one from anywhere in England to a 93.3pp lower chance for the worst CCG. Table 1(b) reports a between-CCG lottery index of 0.2931 based on percentage achievement of the maximum score, with the alternative measure of percentage achievement of points available to the practice yielding the same result to 4 significant figures. Lottery indices for the separate clinical, public health and public health additional services domains are somewhat lower, but all are above 0.2 despite more than half of practices achieving the maximum score in the latter two domains. These considerably higher estimates of the variation in care quality compared to both the GPPS and CQC indices cannot simply be dismissed as an artefact of the cardinality of QOF scores: collapsing the total QOF score into a 5-category variable with population proportions for England as a whole identical to those for the GPPS measure only reduces the index value to 0.2577. Rather they would appear to reflect the relatively high degree of variation in QOF scores between CCGs as compared to within CCGs, with the between-CCG standard deviation of 3.0 pp reported above similar in magnitude to a weighted-average within-CCG standard deviation of practice quality of 3.8 pp: between-CCG differences accounted for as much as 30.4% of the overall variance in practice-level total QOF scores. Figure 7(b) shows that the association between the ranking of CCGs by QOF achievement and GPPS patient experience is weak with the Kendall's τ_a of 0.333 (95% CI, 0.236 to 0.431) implying that QOF scores also fail to provide ordinally equivalent information to the GPPS data on some common latent 'primary care quality' characteristic. Finally, 32.1% of variation

(0.0938/0.2925) in QOF achievement between CCGs was accounted for by differences in the demographic composition of practice lists, with the GLDRM yielding a somewhat higher estimate of the proportion of 'explained' variation in this case. Illustrative distribution regression model results for QOF achievement are presented in the Appendix.

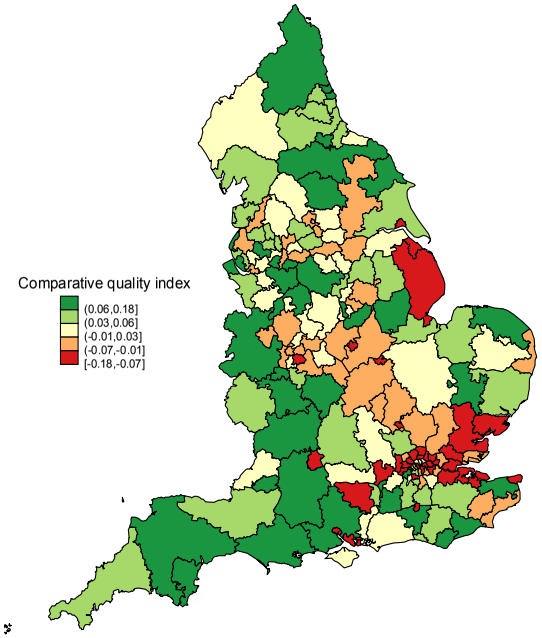
By way of summary, Figure 8 maps the comparative quality indices by CCG quintile for the three alternative practice quality indicators. The maps share some similar features, which is to be expected given the positive association between the corresponding comparative quality indices. In particular, all show a higher concentration of CCGs with poorer levels of primary care quality in the London area. Nevertheless, the prevailing impression is of pervasive differences in the ranking of individual CCGs across the three measures, with nine CCGs in the top quintile on one measure and the bottom on another.

Discussion

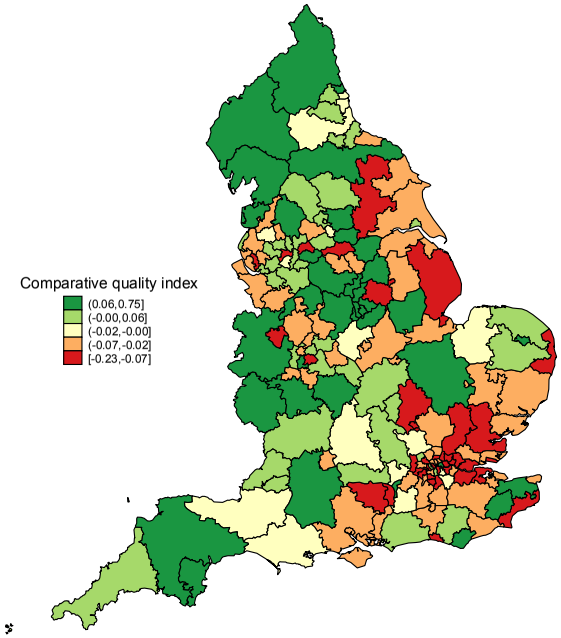
Evidence of the variation in the quality of healthcare services is increasingly being provided by multicategory response information from patient experience surveys, supplementing the routine collection of standard cardinal quality indicators. This paper proposes a measurement framework that is directly applicable to both ordinal and cardinal quality indicators, providing intelligible, patient-oriented measures of both the comparative quality of each member of a set of healthcare providers serving some population and the variation in quality between them. Our approach is motivated by the concept of statistical preference whereby one healthcare provider is judged to be better than another if the patient receiving the (strictly) higher quality care of any randomly chosen pair of patients is more likely to be registered with the first rather than the second provider. Unlike first order stochastic dominance, statistical preference will provide a graded comparison of all possible pairs of care quality profiles. The resultant measures are sensitive to the full distribution

Figure 8. Comparative quality indices by CCG quintile based on alternative primary care quality indicators:

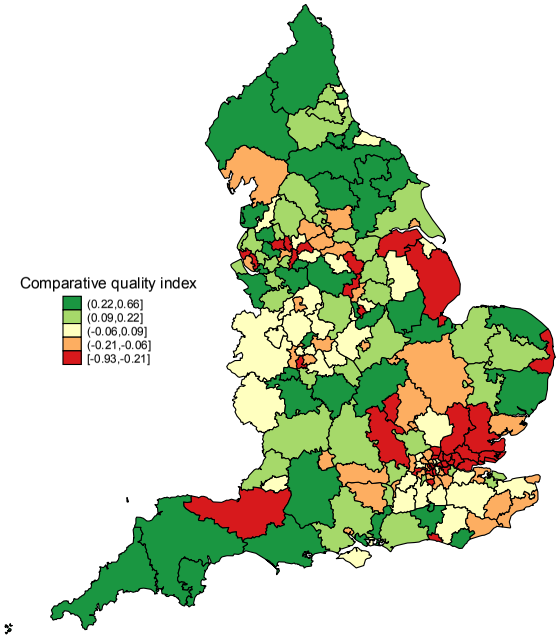
(a) GPPS 5-category indicator



(b) CQC inspection ratings



(c) QOF achievement scores



of quality scores for each provider, not just the mean nor the proportion meeting some binary quality threshold.

The GPPS offers a large-scale, annual survey of patients' experience in virtually all practices in England, with practice-level multicategory response data made publicly available in a timely fashion. We find significant variation in primary care quality levels both between practices within individual CCGs and between CCGs in 2019, with the right to choose between any two practices within a CCG leading on average to an 8.9 pp higher chance that patient experience would be better than it would otherwise have been under random assignment. Dichotomisation leads to a reduction in measured variation, with the loss of contrast most marked between 'average' and 'good' providers. Practice-level information on primary care quality is also available in the form of ordinal CQC inspection ratings and cardinal QOF achievement scores, which are generated for regulatory and performance incentive purposes respectively. We show that neither provide an alternative source of ordinally equivalent information to the GPPS survey on some common latent 'primary care quality' variable. Additionally, the measured level of between-CCG variation is much higher using QOF scores than with the other two quality indicators. Why this is the case is unclear though we do demonstrate that it is not due to the cardinality of the QOF indicator by showing that the value of the lottery index is relatively insensitive to the grouping of QOF scores.

Elimination of the postcode lottery in GP patient experience would provide a measurable, policy-relevant objective to the extent that such variation was due to factors within the control of the National Health Service. In particular, attainment of the goal would not require that all individual patients could expect to receive the same quality of care, which is surely unrealistic, but rather that their experience was equally likely to be better rather than worse as a result of being registered with one practice or CCG rather than another to the extent that this was achievable. Our

findings indicate that patient experience tends to be worse in urban practices and CCGs with higher proportions of prime working age and ethnic minority patients. This might suggest the need for case-mix adjustment of patient experience profiles to ensure equitable comparison among providers but this practice is controversial in that it effectively discounts any differences in response tendencies between different patient subgroups and also runs the risk of 'masking' systematic disparities in the actual standard of care provided to them (see Paddison et al., 2012 for further discussion). We further note that, even if it was thought appropriate to adjust for case-mix, the indirectly standardised profiles generated in this study cannot be used as the basis for such a procedure as the distribution regression models are estimated using practice not patient level data.

In conclusion, the proposed approach provides a general framework to measure variation between healthcare providers or geographical areas making full use of the information provided by the ordinal quality indicators that are now routinely available. Further studies are required to explore whether our empirical findings are more generally characteristic of the scale of healthcare variation in other clinical settings and countries. It would also be of interest to use the framework to track changes in healthcare quality over time, with the impact of the COVID-19 epidemic on levels of GP patient experience an obvious topic for investigation. More work could also be done to explore the determinants of patient experience using individual data to avoid the potential for ecological bias from the use of practice-level data.

References

Allanson, P., 2021. Ordinal health disparities between population subgroups: measurement and multivariate analysis with an application to the North-South divide in England. Journal of Economic Inequality, forthcoming.

Allison, R.A., Forster, J.E., 2004. Measuring health inequality using qualitative data. J. Health Econ. 23(3), 505–524.

Bond T.N., Lang K., 2019. The Sad Truth about Happiness Scales. Journal of Political Economy, 127: 1629- 1640.

Bruyneel L., Tambuyzer E., Coeckelberghs E., Wachter D.D., Sermeus W., Ridder D.D., Ramaekers, D., Weeghmans, I., Vanhaecht, K., 2017. New instrument to measure hospital patient experiences in Flanders. Int J Environ Res Public Health, 14:1319

Busse R., Klazinga N., Panteli D., Quentin W. (Eds.), 2019. Improving healthcare quality in Europe. Characteristics, effectiveness and implementation of different strategies. United Kingdom: World Health Organization (WHO), Organisation for Economic Co-operation and Development (OECD).

Care Quality Commission, 2019a. The state of health care and adult social care in England 2018/19. London: HMSO. Available at: https://www.cqc.org.uk/sites/default/files/20191015b_stateofcare1819_fullreport.pdf

Care Quality Commission, 2019b. Care Directory with ratings (02 January 2019). Available at: https://www.cqc.org.uk/about-us/transparency/using-cqc-data

Chernozhukov, V., Fernández-Val, I., Melly, B., 2013. Inference on counterfactual distributions. Econometrica, 81, 2205–2268.

Dartmouth Atlas of Health Care. Available at: http://www.dartmouthatlas.org/

De Baets, B., De Meyer, H., 2007. Toward Graded and Nongraded Variants of Stochastic Dominance. In I. Batyrshin, L. Sheremetov, L.A. Zadeh (Eds.), Perception-based Data Mining and Decision Making in Economics and Finance, Studies in Computational Intelligence (Volume 36, pp.261–274). Berlin Heidelberg: Springer-Verlag.

DeCourcy, A., West, E., Barron, D., 2102. The National Adult Inpatient Survey conducted in the English National Health Service from 2002 to 2009: how have the data been used and what do we know as a result? BMC Health Serv Res 12, 71, 12pp.

De Schuymer, B., De Meyer, H., De Baets, B., 2003. A fuzzy approach to stochastic dominance of random variables. In: T. Bildi, B. De Baets, O. Kaynak (Eds.), Lecture Notes in Artificial Intelligence (Volume 2715, pp.253–260). Berlin Heidelberg: Springer-Verlag.

Ibáñez, B., Librero, J., Bernal-Delgado, E., Peiró, S., López-Valcarcel, B.G., Martínez, N., Aizpuru, F., 2009. Is there much variation in variation? Revisiting statistics of small area variation in health services research. BMC Health Serv Res 9, 60, 12pp. https://doi.org/10.1186/1472-6963-9-60

Ipsos MORI, 2019. GP Patient Survey 2019: Technical annex. Available at: https://gp-patient.co.uk/surveysandreports2019

Le Breton, M., Michelangeli, A., Peluso, E., 2008. Wage Discrimination Measurement: In Defense of a Simple but Informative Statistical Tool. Università Commerciale Luigi Bocconi Centre for Research on the Public Sector Working Paper No. 112.

Lyratzopoulos, G., Elliot, M., Barbiere, J.M., Henderson, A., Staetsky, L., Paddison, C., Campbell, J., Roland, M., 2012. Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. BMJ Quality and Safety, 21: 21-29.

Mainz, J., 2003. Defining and classifying clinical indicators for quality improvement. International Journal for Quality in Health Care 15(6), 523–530.

Monti, M., Santori, A., 2011. Stratification and between-group inequality: A new interpretation. Review of Income and Wealth 57 (3), 412–427.

NHS Digital, 2018. Patients registered at a GP Practice - December 2018. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice/december-2018

NHS Digital, 2019. Quality and Outcomes Framework, Achievement, prevalence and exceptions data 2018-19 [PAS]. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data

NHS England, 2019. The GP Patient Survey: Practice report (2019 publication). Available at: https://gp-patient.co.uk/surveysandreports2019

NHS England and Ipsos MORI, 2019. GP Patient Survey: National report 2019. Available at: https://gp-patient.co.uk/surveysandreports2019

NHS England, 2021. National Patient and Staff Surveys. Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/patient-surveys/

Paddison, C., Elliot, M., Parker R., Staetsky, L., Lyratzopoulos, G., Campbell, J.L., Roland, M., 2102. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. BMJ Quality and Safety, 21: 634-40.

Public Health England, 2015. NHS Atlas of Variation in Healthcare, Reducing unwarranted variation to increase value and improve quality. September. PHE Publications.

Pyatt, G., 1976. On the interpretation and disaggregation of Gini coefficients. The Economic Journal 86 (342), 243-255.

Santos, R., Gravelle, H., Propper, C., 2017. Does quality affect patients' choice of doctor? Evidence from England. The Economic Journal, 127 (March), 445–494.

Yitzhaki, S., Lerman, R., 1991. Income stratification and income inequality. Review of Income and Wealth 37 (3), 313-329.

Appendix. Selected distribution regression model results

Table A1. Fixed effects estimates of GLDRM for GPPS 5-category indicator (full GPPS sample)

	Reported experience no better than:			
	very poor	fairly poor	neither good nor poor	fairly good
	$P(q\leq1)$	$P(q\leq2)$	$P(q\leq3)$	$I(q\leq4)$
FEMALE	-0.2321	-0.0915	-0.2169*	-0.0482
	0.1381	*0.1259*	*0.1048*	*0.0860*
AGE 16-24	-0.8837**	-0.6475**	-0.2461	-0.0839
	0.2068	*0.1740*	*0.1448*	*0.1320*
AGE 25-34	-0.3410	-0.2635	-0.0420	0.1972
	0.1954	*0.1630*	*0.1410*	*0.1257*
AGE 35-44	-0.3300	-0.2092	-0.0289	0.2299
	0.2136	*0.1757*	*0.1458*	*0.1375*
AGE 55-64	-0.5117*	-0.4240*	-0.3053	-0.0805
	0.2469	*0.2060*	*0.1782*	*0.1678*
AGE 65-74	-1.0596**	-1.0555**	-0.9574**	-0.7796**
	0.3499	*0.2817*	*0.2433*	*0.2162*
AGE 75-84	-1.2870**	-1.2519**	-0.6516	-0.4369
	0.4554	*0.4371*	*0.3418*	*0.2769*
AGE 85+	-1.4187*	-1.1831*	-1.1785**	-0.8677*
	0.6616	*0.5490*	*0.4341*	*0.3983*
BLACK	0.4737	0.0764	0.2241	0.3427
	0.4031	*0.3191*	*0.2777*	*0.2675*
ASIAN	0.5480**	0.4118**	0.4246**	0.4700**
	0.0558	*0.0542*	*0.0503*	*0.0499*
MIXED	0.4711**	0.3070**	0.3308**	0.3602**
	0.1312	*0.1027*	*0.1140*	*0.1118*
OTHER	0.3937	0.3015	0.3616	0.3796*
	0.2502	*0.1999*	*0.1881*	*0.1903*
constant	-1.5291**	-1.1609**	-0.7021**	0.0590
	0.1548	*0.1310*	*0.1501*	*0.1010*
Practices	6926	6926	6926	6926
CCG clusters	195	195	195	195

*The dependent variable $P(q\leq c)$ takes a value equal to the proportion of patients in a practice reporting their experience as no better than category c ($c=1,2,3,4$). CCG fixed effects not reported. Semirobust CCG-clustered standard errors are in italics. *$p<0.05$, **$p<0.01$. Source: Own calculations from GPPS data.*

Table A2. Fixed effects estimates of the LPDRM: CQC 4-category (matched sample)

	Inspection rating no better than:		
	inadequate	requires improvement	good
	P(q≤1)	*P(q≤2)*	*P(q≤3)*
FEMALE	-0.0018	-0.0468	-0.1601*
	0.0266	*0.0432*	*0.0783*
AGE 16-24	0.0271	0.0405	-0.1505
	0.0398	*0.0733*	*0.1748*
AGE 25-34	-0.0103	0.0508	-0.1574
	0.0325	*0.0746*	*0.1164*
AGE 35-44	-0.0087	-0.0014	-0.1094
	0.0346	*0.0756*	*0.1229*
AGE 55-64	-0.0430	0.0584	0.1546
	0.0360	*0.0938*	*0.1390*
AGE 65-74	-0.0090	-0.1450	0.0641
	0.0446	*0.1275*	*0.1608*
AGE 75-84	0.0046	0.0406	-0.1896
	0.0528	*0.1270*	*0.1969*
AGE 85+	0.1275	0.3043	-0.4258
	0.1147	*0.2079*	*0.3665*
BLACK	0.0167	0.1264*	-0.1042
	0.0230	*0.0598*	*0.0683*
ASIAN	0.0267	0.0761*	0.0777*
	0.0191	*0.0328*	*0.0345*
MIXED	0.0072	-0.1794	0.2326
	0.0700	*0.1255*	*0.2281*
OTHER	0.0387	0.0173	0.1898
	0.0426	*0.0916*	*0.0975*
constant	0.0042	0.0079	1.0524**
	0.0273	*0.0592*	*0.0912*
Practices	6926	6926	6926
CCG clusters	195	195	195
R^2	0.044	0.059	0.116
RMSE	0.084	0.183	0.228

*The dependent variable P(q≤c) takes a value equal to takes a value of one if practice quality is no better than category c (c=1,2,3) and zero otherwise. Positive (negative) coefficients imply higher (lower) chances than for the reference group of White men aged 45-54 in NHS Darlington CCG: for example, a 1pp increase in the proportion of Asian patients is predicted to lead to ceteris paribus increases of 0.0267pp, 0.0761pp, and 0.0777pp in the chances of an inspection rating no better than inadequate, requires improvement, and good respectively. CCG fixed effects not reported. Robust CCG-clustered standard errors are in italics. *p<0.05, **p<0.01. Source: Own calculations from CQC data.*

Table A3. Fixed effects estimates of the GLDRM: CQC 4-category (matched sample)

	Inspection rating no better than:		
	inadequate	requires improvement	good
	P(q≤1)	*P(q≤2)*	*P(q≤3)*
FEMALE	-0.4068	-0.6023	-1.4906*
	1.1849	*0.5980*	*0.6732*
AGE 16-24	0.7740	0.5432	-1.0384
	1.6620	*0.9706*	*1.2663*
AGE 25-34	-0.9021	0.7694	-1.5232
	1.9052	*1.0279*	*1.1469*
AGE 35-44	-0.6966	-0.0028	-1.5676
	2.0700	*1.0176*	*1.2995*
AGE 55-64	-3.5689	0.9098	1.7693
	2.4661	*1.3396*	*1.5318*
AGE 65-74	-0.5865	-1.9342	0.7764
	2.8765	*1.8436*	*1.7287*
AGE 75-84	0.0530	0.5705	-2.1685
	3.5538	*1.9501*	*1.9666*
AGE 85+	7.8997	4.7724	-3.7078
	7.2872	*3.1906*	*3.1711*
BLACK	1.0675	1.3407**	-1.8428
	0.7728	*0.5178*	*1.0421*
ASIAN	1.0573*	0.7611*	1.1880*
	0.5377	*0.3032*	*0.5018*
MIXED	1.2266	-1.8981	1.5062
	2.8351	*1.5183*	*2.4703*
OTHER	2.0818	0.5158	2.9806
	1.1754	*0.8527*	*1.5861*
constant	-4.3216**	-4.7979**	2.6138**
	1.4483	*0.8574*	*0.8810*
Practices	6926	6926	6926
CCG clusters	195	195	195

*The dependent variable P(q≤c) takes a value equal to takes a value of one if practice quality is no better than category c (c=1,2,3) and zero otherwise. CCG fixed effects not reported. Semirobust CCG-clustered standard errors are in italics. *p<0.05, **p<0.01. Source: Own calculations from CQC data.*

Table A4. Fixed effects estimates of the LPDRM: 5-category grouping of QOF score (matched sample)

	Reported experience no better than:			
	Category1 $P(q\leq1)$	Category2 $P(q\leq2)$	Category3 $P(q\leq3)$	Category4 $I(q\leq4)$
FEMALE	0.0705	-0.0815	-0.1011	-0.1483
	0.0475	*0.0741*	*0.1112*	*0.1306*
AGE 16-24	0.3776*	0.4543**	0.5647**	0.3147
	0.1483	*0.1572*	*0.1652*	*0.1748*
AGE 25-34	0.0619	0.0507	0.1116	0.2394
	0.0471	*0.0950*	*0.1529*	*0.1754*
AGE 35-44	0.0724	0.1670	0.0023	0.0596
	0.0459	*0.1014*	*0.1444*	*0.2279*
AGE 55-64	-0.0314	0.0045	-0.0788	-0.0187
	0.0611	*0.1340*	*0.1534*	*0.2636*
AGE 65-74	0.0492	0.0697	0.0041	0.4457
	0.1089	*0.1547*	*0.2415*	*0.3074*
AGE 75-84	0.0224	-0.0759	-0.4044	-0.9528**
	0.0713	*0.1934*	*0.3048*	*0.3536*
AGE 85+	-0.0627	-0.0715	0.2149	-0.2980
	0.1010	*0.2345*	*0.4011*	*0.5129*
BLACK	0.0505	0.1652*	0.1672	0.6050**
	0.0459	*0.0812*	*0.1530*	*0.1622*
ASIAN	-0.0296	-0.0737	-0.0509	0.1287
	0.0220	*0.0383*	*0.0579*	*0.0729*
MIXED	0.4814	0.4426	0.7102*	1.1398**
	0.3184	*0.3356*	*0.3096*	*0.3984*
OTHER	-0.0852	0.0326	0.1230	0.1240
	0.0606	*0.1492*	*0.1972*	*0.2409*
constant	-0.0941	-0.0354	0.0184	0.2556
	0.0500	*0.0800*	*0.1019*	*0.1524*
Practices	6926	6926	6926	6926
CCG clusters	195	195	195	195
R^2	0.317	0.256	0.180	0.158
RMSE	0.118	0.215	0.346	0.464

*The dependent variable P(q≤c) takes a value equal to takes a value of one if practice quality is no better than category c (c=1,2,3,4) and zero otherwise. Grouping of data into categories based on population proportions for GPPS 5-category indicator. Positive (negative) coefficients imply higher (lower) chances than for the reference group of White men aged 45-54 in NHS Darlington CCG: for example, a 1 pp increase in the proportion of Asian patients is predicted to lead to ceteris paribus reductions of 0.0296pp, 0.0737pp, and 0.0509pp in the chances of a QOF score no better than category 1, 2 and 3 respectively, and an increase of 0.1287pp in the probability of a score no better than category 4. CCG fixed effects not reported. Robust CCG-clustered standard errors are in italics. *p<0.05, **p<0.01. Source: Own calculations from QOF data.*

Table A5. Fixed effects estimates of the GLDRM: 5-category grouping of QOF score (matched sample)

	Reported experience no better than:			
	Category1	Category2	Category3	Category4
	P(q≤1)	*P(q≤2)*	*P(q≤3)*	*I(q≤4)*
FEMALE	-0.0464	-1.3087*	-0.4983	-0.3852
	0.8455	*0.6294*	*0.4765*	*0.3730*
AGE 16-24	5.3566**	3.1483**	1.8557**	0.9840
	1.4002	*0.9805*	*0.6072*	*0.5350*
AGE 25-34	0.4622	0.0802	0.2461	0.7040
	1.5671	*0.9602*	*0.6379*	*0.5137*
AGE 35-44	2.9795	1.5907	-0.0414	0.1417
	1.9228	*1.0600*	*0.6609*	*0.6683*
AGE 55-64	-1.7207	0.2368	-0.3693	-0.0330
	2.7701	*1.4848*	*0.7232*	*0.7569*
AGE 65-74	1.0903	0.5770	-0.2058	1.3389
	4.5302	*1.8130*	*1.1575*	*0.8768*
AGE 75-84	-0.4889	-1.6843	-2.3955	-2.7368**
	2.9302	*2.2305*	*1.4959*	*1.0235*
AGE 85+	-3.7233	-1.3042	0.9713	-0.9489
	3.7063	*2.9710*	*2.0828*	*1.4702*
BLACK	2.1150*	1.2580*	0.6041	1.8963**
	1.0013	*0.5585*	*0.5497*	*0.5103*
ASIAN	-0.5662	-0.5421	-0.1432	0.3411
	0.6814	*0.3621*	*0.2266*	*0.2122*
MIXED	5.0856	2.3526	2.5038*	3.6978**
	2.8055	*1.8146*	*1.1117*	*1.3408*
OTHER	-0.8971	0.2253	0.4655	0.3309
	1.9566	*1.0139*	*0.6930*	*0.7210*
constant	-5.8428**	-4.6817**	-4.3388**	-0.6965
	1.3176	*0.7983*	*0.4983*	*0.4404*
Practices	6926	6926	6926	6926
CCG clusters	195	195	195	195

*The dependent variable P(q≤c) takes a value equal to takes a value of one if practice quality is no better than category c (c=1,2,3,4) and zero otherwise. Grouping of data into categories based on population proportions for GPPS 5-category indicator. CCG fixed effects not reported. Semirobust CCG-clustered standard errors are in italics. *p<0.05, **p<0.01. Source: Own calculations from QOF data.*